Make Chicago "Safe" Again

By: Tyler Mooney, Garrett Glissmann, Luis Velzte and Ryan Close

Description

The goal of this project is to find interesting trends using the city of Chicago's crime reports that spans the years 2001 to early 2018. We would perform a longitudinal survey of crime trends. We hope to find geographical patterns as well as discover connections to other public health factors (e.g. socioeconomic indicators, birth rates, etc.).

Prior work

ANALYSIS OF CRIMES IN CHICAGO 2001 - 2017 - *VIVEK MANGIPUDI*

https://rstudio-pubs-static.s3.amazonaws.com/294927_b602318d06b74e4cb2e6be336522e94e.html

Crime Over Time: Visualizing Crime Data in Chicago - *Greg Laughlin*

https://socrata.com/blog/crime-time-visualizing-crime-data-chicago/

48 YEARS OF CRIME IN CHICAGO: A Descriptive Analysis of Serious Crime Trends from 1965 to 20131 (Andrew V. Papachristos)

https://isps.yale.edu/sites/default/files/publication/2013/12/48yearsofcrime_final_ispsworkingpaper023.pdf

Data sets

The datasets come from the city of Chicago's Data Catalog (https://data.cityofchicago.org/)

Primary dataset of crime reports: "Crimes 2001-Present"

Other datasets (possible):

- "Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012"
- "Public Health Statistics - Births and birth rates in Chicago, by year, 1999 – 2009"
- "Affordable Housing Units by Community Area"
- "Public Health Statistics - Prenatal care in Chicago, by year, 1999 – 2009"

Proposed Work

First, we will need to preprocess the primary dataset (crime reports). This includes ensuring all the attributes are correctly formatted, such as date fields, as well as filling in or removing missing data. If we use the other datasets those will need to be preprocessed too. The challenge is that not all the datasets are longitudinal or they represent a smaller time range. So we will need to adjust for that difference. Our main goal is then to group the data into time segments and then find patterns by geographical region. We can find simple metrics such as average crime rate, median, and standard deviation using Pandas in Python. However, we hope to go beyond this and use machine learning / data mining tools (e.g. Scipy, TensorFlow, WEKA, RapidMiner) to find more interesting data patterns.

List of Tools

- Jupyter Notebook
  - Heat Mapping (folium)
- RapidMiner
  - Data Preprocessing and Visualization
  - Predictive Analytics and Statistical Modeling
  - Evaluation
- Weka
  - Data Preprocessing, Clustering, Classification
  - Data Regression, Visualization and Feature Selection
- Scipy / Tensorflow

Evaluation

Firstly, we hope to arrive at comparable results to previous studies of Chicago's crime. Secondly, we want to go further than previous studies and find novel trends between crime and other social and public factors. This knowledge could potentially influence Chicago's crime reduction strategies. For example, if we found that affordable housing was correlated with lower rates of crime that would be valuable knowledge for the Chicago city planners.